

A Data-Driven Compression Method for Transient Rendering

Yun Liang
South China Agricultural University

Mingqin Chen
South China Agricultural University

Zesheng Huang
South China Agricultural University

Diego Gutierrez
Universidad de Zaragoza, I3A

Adolfo Muñoz
Universidad de Zaragoza, I3A

Julio Marco
Universidad de Zaragoza, I3A
juliom@unizar.es

ABSTRACT

Monte Carlo methods for transient rendering have become a powerful instrument to generate reliable data in transient imaging applications, either for benchmarking, analysis, or as a source for data-driven approaches. However, due to the increased dimensionality of time-resolved renders, storage and data bandwidth are significant limiting constraints, where a single time-resolved render of a scene can take several hundreds of megabytes. In this work we propose a learning-based approach that makes use of deep encoder-decoder architectures to learn lower-dimensional feature vectors of time-resolved pixels. We demonstrate how our method is capable of compressing transient renders up to a factor of 32, and recover the full transient profile making use of a decoder. Additionally, we show how our learned features significantly mitigate variance on the recovered signal, addressing one of the pathological problems in transient rendering.

CCS CONCEPTS

•Computing methodologies → Ray tracing;

KEYWORDS

Transient rendering, compression, denoising

ACM Reference format:

Yun Liang, Mingqin Chen, Zesheng Huang, Diego Gutierrez, Adolfo Muñoz, and Julio Marco. 2019. A Data-Driven Compression Method for Transient Rendering. In *Proceedings of SIGGRAPH '19 Posters, Los Angeles, CA, USA, July 28 - August 01, 2019*, 2 pages.
DOI: 10.1145/3306214.3338582

1 INTRODUCTION

The recent advances on transient imaging have enabled a wide variety of applications such as non-line-of-sight (NLOS), material classification, or visualization of light in motion [Jarabo et al. 2017]. Recently, a novel method has been presented yielding a new class of imaging algorithms mimicking the various capabilities of LOS cameras [Liu et al. 2019]. Access to reliable time-resolved light transport has therefore become a valuable asset in transient imaging, not only for benchmarking and analysis, but also as a source of reference data in data-driven methods. Monte Carlo methods for transient rendering [Jarabo et al. 2014; Marco et al. 2019] allow

us to synthesize reliable time-resolved light transport, and have already been proved useful in data-driven methods that address transient imaging problems [Guo et al. 2018; Marco et al. 2017]. However, the increased dimensionality of time-resolved renders dramatically boosts the size of the data, proportional to the resolution of the temporal dimension. This is a critical factor when storage or data bandwidth are limiting constraints of a transient imaging system, with input/output operations becoming a potential bottleneck during data processing.

To address these issues, in this work we leverage the properties of encoder-decoder neural networks to extract a reduced set of features from the radiance information available during the rendering step. We propose a fully-convolutional architecture that allows to reduce temporal profiles of arbitrary resolution by a factor of 32, and we benefit from the denoising properties of encoder-decoders [Vincent et al. 2008] to mitigate variance in the decompressed signal.

2 OUR APPROACH

Our goal is to use the radiance information generated during rendering to extract a reduced set of features f_L that will represent a *compressed* radiance profile $L_c(t)$, i.e. a single time-resolved pixel. This set of features should allow us to recover the original radiance profile in a decompression step. Since a transient render can have arbitrary temporal resolution, it is of interest that our approach is capable of applying the same compression ratio to temporal profiles of any length.

To meet these requirements, we rely on a fully-convolutional encoder-decoder architecture, where the encoder is in charge of extracting meaningful features from the rendering data (compression step), and the decoder uses these to recover the original time-resolved pixel L_c (decompression step). To maximize the information available during rendering, the encoder applies spatio-temporal convolutions over a neighborhood $L_{9 \times 9}$ of size $9 \times 9 \times T$ centered at the pixel L_c to be compressed, where T is any arbitrary temporal resolution. By applying subsequent 3D convolution operations with non-linear activations, the encoder extracts a feature vector f_L of size $T/32$. The decoder then applies multiple 1D deconvolutions and convolutions in order to recover the time-resolved pixel L_c . Following previous works on deep residual nets [He et al. 2016] we include residual connections in the decoder. Our convolutional blocks are composed by a convolution layer, batch normalization layer, and a tangent activation (with the exception of a rectifier linear unit in the output layer). A detailed description of our network can be found in Figure 1.

Dataset. As a labeled data source for training and validation, we rely on the publicly available *Zaragoza-DeepToF* transient dataset

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGGRAPH '19 Posters, Los Angeles, CA, USA

© 2019 Copyright held by the owner/author(s). 978-1-4503-6314-3/19/07...\$15.00
DOI: 10.1145/3306214.3338582

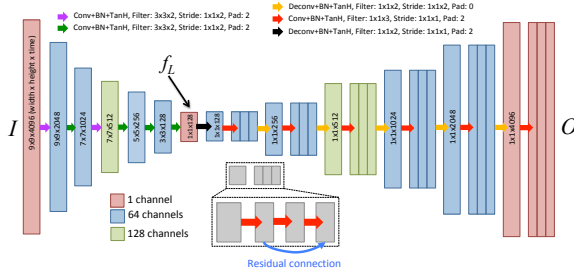


Figure 1: Our proposed architecture, with example sizes for a temporal resolution of $T = 4096$. The encoder extracts a total of $T/32$ features f_L from a $9 \times 9 \times T$ spatial neighborhood, centered at the time-resolved pixel L_c to compress. The decoder uses these features to decompress the time-resolved pixel of size $1 \times 1 \times T$.

[Marco et al. 2017]. This dataset contains time-resolved renders for a wide variety of complex architectural scenarios with a temporal resolution of 4096 pixels at 1.6 picoseconds/pixel. We randomly select a total of 1015000 inputs of size $9 \times 9 \times 4096$ from 145 scenes for training, and a total of 370000 inputs from 37 completely different scenes for validation.

Data preprocessing. Due to attenuation of radiance at every light bounce on the surfaces, time-resolved radiance profiles tend to present exponential decay over time. To ensure that our approach is able to extract features on the whole temporal profile without neglecting attenuated radiance, we perform a logarithmic transformation over the radiance L , defining the input I as

$$I = \begin{cases} \log(L) + 7 & L \geq 1e-7 \\ 0 & L < 1e-7 \end{cases} \quad (1)$$

Loss function. We minimize the error between the reference and the decompressed time-resolved radiance using a mean squared error metric. Since our network operates over the logarithmic transformation of radiance L (Equation 1), we apply an exponential operation over the network output O to compute the mean squared error on linear space, having

$$Loss = \frac{1}{T} \sum_{t=1}^T (b^{I_c(t)} - b^{O(t)})^2, \quad (2)$$

where b is the basis of the exponential function. In practice, we found that choosing $b = 2$ provides a good tradeoff for recovering both peaks and indirect bounces (see Figure 2).

As observed in the validation results of Figure 2, our method is capable of compressing transient renders by a factor of 32 (from 4096 pixels to 128 features), while successfully decompressing the full temporal profile back to 4096 pixels preserving both direct (initial peak) and indirect illumination features. In Figure 3 we show reference frames of the *House* scene from the validation set, and the resulting frames after compressing each reference time-resolved pixel to 128 features and decompressing them back to 4096 pixels. Our decoder is capable of recovering both direct and indirect illumination, while performing significant denoising due to the spatio-temporal features learned by the encoder.

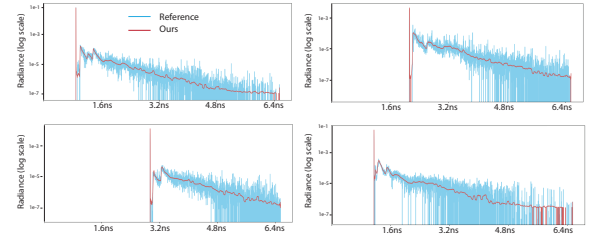


Figure 2: Results for our validation set, showing time-resolved profiles of size 4096 (reference, blue), compressed to 128 features and later decompressed back to 4096 pixels (ours, red). We can observe how our network is able to successfully recover both the initial peak and the indirect illumination with attenuated radiance.

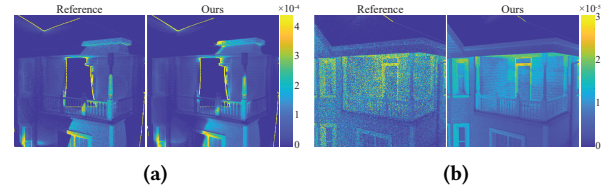


Figure 3: Reference and decompressed frames for the scene *House* from the validation set. Our method successfully decompresses both (a) early frames with direct illumination, and (b) later ones with indirect illumination, while significantly mitigating variance.

ACKNOWLEDGMENTS

This project was funded by DARPA (project REVEAL), the European Research Council under the EU's Horizon 2020 research and innovation programme (project CHAMELEON, grant No. 682080), the Spanish Ministry of Economy and Competitiveness (project TIN2016-78753-P), the National Natural Science Fund of China (61772209), and the Science and Technology Planning Project of Guangdong Province (2016A050502050).

REFERENCES

- Qi Guo, Iuri Frosio, Orazio Gallo, Todd Zickler, and Jan Kautz. 2018. Tackling 3D ToF Artifacts Through Learning and the FLAT Dataset. In *The European Conference on Computer Vision (ECCV)*.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE CVPR*. 770–778.
- Adrian Jarabo, Julio Marco, Adolfo Muñoz, Raul Buisan, Wojciech Jarosz, and Diego Gutierrez. 2014. A Framework for Transient Rendering. *ACM Trans. Graph.* 33, 6 (2014).
- Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. 2017. Recent Advances in Transient Imaging: A Computer Graphics and Vision Perspective. *Visual Informatics* 1, 1 (2017).
- Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. 2019. Non-Line-of-Sight Imaging using Phasor Fields Virtual Wave Optics. *Nature* (2019).
- Julio Marco, Ibón Guillén, Wojciech Jarosz, Diego Gutierrez, and Adrian Jarabo. 2019. Progressive Transient Photon Beams. *Computer Graphics Forum* (2019).
- Julio Marco, Quercus Hernandez, Adolfo Muñoz, Yue Dong, Adrian Jarabo, Min Kim, Xin Tong, and Diego Gutierrez. 2017. DeepToF: Off-the-Shelf Real-Time Correction of Multipath Interference in Time-of-Flight Imaging. *ACM Trans. Graph.* 36, 6, Article 219 (2017).
- Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. ACM, 1096–1103.